REMARKS

Upon entry of the amendments, claims 1-12, 17, 22-25, 33-37, 39, 56 constitute the pending claims in the present application.

Claims 1, 3-7, 9-10, 17, and 24 have been amended for greater clarity. Support for the claim amendments can be found throughout the specification (e.g., page 15, lines 13-16; page 19, lines 31-32; and page 20, lines 1-2). No new matter has been introduced and no new issue has been raised. The amendments are made solely to expedite prosecution of the application, and Applicants reserve the right to prosecute claims of similar or differing scope in subsequent applications.

Applicants respectfully request reconsideration in view of the following remarks. Issues raised by the Examiner will be addressed below in the order they appear in the prior Office Action.

Election/Restriction

The Examiner has acknowledged Applicants' election, with traverse, of Group I (claims 1-12, 17, and 22-25) in the Response filed on October 5, 2008. The Examiner has also acknowledged Applicants' election of three species (TRCP1, TNFR1, and TRAP2) in the Responses filed on October 5, 2007 and March 11, 2008.

However, the Examiner asserts that "claims 3, 4, 6, 7, 9, 10, 24, 33-37, 39, and 44-56 are withdrawn from further consideration pursuant to 37 CFR 1.142(b), as being drawn to nonelected inventions and species, there being no allowable generic or linking claim." See Office Action, page 4, lines 14-17.

Applicants respectfully disagree for the following reasons.

First of all, Applicants remind the Examiner that original claim 1 is drawn to a protein complex comprising . . . (iii) "**at least one** polypeptide **selected from the group** consisting of: NAK, RasGAP3, TRCP1, TRCP2, and a functional variant thereof" (emphasis added). One of skill in the art would appreciate that the claimed protein complex may optionally comprise two, three, or four polypeptides selected from NAK, RasGAP3, TRCP1, and TRCP2.

In both Responses filed on October 5, 2007 and March 11, 2008, Applicants elected TRCP1, with traverse, as a species for search purpose only. Since original claim 1 clearly recites the option of further comprising the other three non-elected species (e.g., NAK, RasGAP3, and TRCP2), dependent claim 10 properly depends from claim 1, although it recites non-elected species in addition to the elected TRCP1 species. In addition, Applicants note that claim 10 recites another elected species (TRAP2). Claim 10 depends from claim 8 which recites the option of further comprising **at least one** polypeptide **selected from**: TRADD, TRAF2, and TRAP2. Thus, dependent claim 10 properly depends from claim 8, although it recites non-elected species in addition to the elected TRAP2 species. Since claim 10 clearly reads on the elected TRCP1 and TRAP2 species, Applicants respectfully request that claim 10 be rejoined into the elected invention.

Similarly, claim 24 properly depends from claims 17 and 23, although it recites non-elected species in addition to the elected TRCP1 and TRAP2 species. Since claim 24 clearly reads on the elected TRCP1 and TRAP2 species, Applicants respectfully request that claim 24 be rejoined into the elected invention.

Furthermore, solely for greater clarity, Applicants have amended withdrawn claims 3-4 and 6-9 to clarify that the protein complex comprises TRCP1 (the elected species) in addition to non-elected species. As such, amended claims 3-4 and 6-9 read on the elected TRCP1 species. Applicants respectfully request that claims 3-4 and 6-9 be rejoined into the elected invention.

Information Disclosure Statement

Applicants note that the Examiner has considered and initialed the Information Disclosure Statements filed on June 13, 2005 and January 23, 2006.

Claim Rejections under 35 U.S.C. § 112, First Paragraph

Claims 1-2, 5, 8, 11-12, 17, 22-23, and 25 are rejected under 35 U.S.C. § 112, first paragraph, as allegedly failing to comply with the written description requirement. Applicants respectfully traverse this rejection to the extent it is maintained over the claims as amended.

Specifically, the Office Action asserts that "[b]ecause the claims are drawn to functional variants of all of the recited polypeptides, these are genus claims . . . With the exception of the wildtype polypeptides explicitly claimed, the skilled artisan cannot envision the detailed

chemical structure of the encompassed polypeptides, and therefore conception is not achieved until reduction to practice has occurred, regardless of the complexity or simplicity of the method of isolation." See Office Action, page 6, lines 1-22.

Applicants respectfully disagree. The specification sufficiently describes the claimed invention. In particular, the term "functional variant" is well defined in the specification. For example, the specification teaches that "[a] 'variant' of a polypeptide, such as, for example, a variant of a TNF-α, a TNFR, a TRCP1 or a TRCP2 includes chimeric proteins, fusion proteins, mutant proteins, proteins having *similar but non-identical sequences*, protein fragments, mimetics, etc, so long as the variant has at least a portion of an amino acid sequence of a native protein, or at least a portion of an amino acid sequence of substantial sequence identity to the native protein. A 'functional variant' includes a variant that *retains at least one function of the native protein*" (e.g., page 15, lines 13-16, emphasis added).

Nevertheless, solely to expedite prosecution of the application, Applicants have amended claims 1, 3, and 17 to clarify the claimed subject matter. Support can be found throughout the specification. For example, the specification discloses that "[i]n other embodiments, the variant polypeptide has an amino acid sequence at least 80%, 85%, 90%, 95%, 97%, 98%, 99% or 100% identical to an amino acid sequence as set forth in any of SEQ ID Nos: 1, 3, 5, 7, 9, 11, 13, 15, 17, and 19" (page 19, lines 31-32; and page 20, lines 1-2). Applicants submit that the claims as amended satisfy the written description requirement.

Applicants wish to draw the Examiner's attention to a recent PTO Board decision, which supports Applicants' position that the specification provides adequate written description for the recited genus of polypeptides and functional variants in the amended claims. *See Ex parte Bandman*, No. 2004-2319, (BPAI 2005).

In *Bandman* (U.S. Application No. 09/915,694), Applicants appealed a Final rejection by the Examiner, and the Board reversed the rejections based on both the written description and enablement requirements of 35 U.S.C. § 112, first paragraph to one of the claims on appeal. The Board found that claims directed to a naturally occurring amino acid (or polynucleotide) sequence at least 95% identical to the disclosed amino acid (or polynucleotide) sequence were enabled and met the written description requirement, even in the absence of explicitly reciting a *functional requirement* of the claimed sequences. The Board noted that "[t]he written

description requirement . . . does not require a description of the complete structure of every species within a chemical genus." *Bandman*, No. 2004-2319 at p. 3. The Board also compared the circumstances of *Bandman* with those faced by the Federal Circuit in *Enzo Biochem, Inc. v. Gen-Probe Inc.*, 296 F.3d 1316 (Fed. Cir. 2002). In *Enzo Biochem*, the Federal Circuit determined that an "[a]dequate written description may be present for a genus of nucleic acids based on their hybridization properties, 'if they hybridize under highly stringent conditions to known sequences because such conditions dictate that all species within the genus will be structurally similar'" (citing *Enzo Biochem*, 296 F.3d at 1324).

Furthermore, Applicants point out that functions of some components of the protein complex (e.g., TNF-α, TNFR, and NAK) were known in the art long before the filing of the subject application (see, e.g., Heyninck et al., 2001, Mol Cell Biol Res Commun, 4:259-65, cited in the Office Action). Also, with respect to the recited polypeptide genus, a number of species representative of the entire genus had been reduced to practice and were generally known in the art at the time the present application was filed (see, e.g., US Patent Nos. 5,028,420, 5,160,483, 5,606,023, 5,773,582, published US Applicant No. US20040170975, and published PCT application WO 01/64889). In fact, "[w]hat is conventional or well known to one of ordinary skill in the art need not be disclosed in detail." *See, e.g., Hybritech, Inc. v. Monoclonal Antibodies, Inc.*, 802 F.2d 1367, 1384, 231 USPQ 81, 94 (Fed. Cir. 1986). In this case, one of skill in the art would have readily understood that the inventor was in possession of these polypeptides as recited in the pending claims, in view of the teachings of the specification (e.g., page 15, lines 13-16; and page 19, lines 31-32; and page 20, lines 1-2) and the knowledge in the art. This alone provides adequate written description for the recited polypeptide genus. "If a skilled artisan would have understood the inventor to be in possession of the claimed invention at the time of filing, even if every nuance of the claims is not explicitly described in the specification, then the adequate description requirement is met. *See, e.g., Vas-Cath*, 935 F.2d at 1563, 19 USPQ2d at 1116; *Martin v. Johnson*, 454 F.2d 746, 751, 172 USPQ 391, 395 (CCPA 1972)." MPEP 2163.

Therefore, the genus as recited in the amended claims is adequately represented by the polypeptides disclosed in the specification and related polypeptides generally known in the art. Applicants' position is supported by MPEP 2163 and the Board decision in *Ex parte Bandman*.

Thus, all pending claims meet the requirement of 35 U.S.C. § 112, first paragraph. Reconsideration and withdrawal of the rejections are respectfully requested.

Claim Rejection under 35 U.S.C. § 112, First Paragraph

Claims 1-2, 5, 8, 11-12, 17, 22-23, and 25 are rejected for lack of enablement. Applicants respectfully traverse these rejections to the extent it is maintained over the claims as amended.

Specifically, the Office Action asserts that "the specification, while being enabling for an isolated, purified, or recombinant protein complex comprising: (i) a tumor necrosis factor alpha (TNF-α) polypeptide; (ii) a TNF-α receptor (TNFR) polypeptide; and (iii) at least one polypeptide selected from: NF-κB activating kinase (NAK), RasGAP3, TRCP1, and TRCP2, does not reasonably provide enablement for an isolated, purified, or recombinant protein complex comprising functional variants of the polypeptides mentioned above." Office Action, page 7, last paragraph.

Applicants respectfully disagree and contend that the specification as filed is enabling for the full scope of the claimed invention. Nonetheless, as mentioned above, Applicants have amended claims 1, 3, and 17 to specify the structural and functional features of the recited polypeptides and their functional variants (e.g., TNF-α, TNFR, NAK, RasGAP3, TRCP1, and TRCP2) which are present in the claimed protein complex. Applicants believe that such amendments render the rejection moot.

As described above, the specification teaches the structural and functional properties of the various polypeptides of the claimed protein complex. In addition, the specification teaches how to make functional variants of those polypeptides without changing the activities (e.g., pages 20-24). For example, the specification describes how to make functional variants of the polypeptides which retain the activity to bind to an interactive polypeptide (e.g., the paragraph bridging pages 23 and 24). Further, a skilled artisan could practice the present invention without necessarily knowing which amino acid substitutions, deletions, or insertions to make, since each of the polypeptides is defined structurally and functionally. The techniques of combinatorial mutagenesis (Reidhaar-Olson and Sauer 1988, enclosed herewith as **Exhibit A**) and high through-put screening, known in the art at the time of filing, make the identification of functional polypeptide variants routine. The fields of combinatorial and scanning mutagenesis

had trivialized the once-complex process of making and testing polypeptide variants long before the filing of the present application. These techniques have been routinely practiced and allow a wide range of amino acid substitutions to be made and tested for the maintenance or disruption of functional properties without undue experimentation.

In sum, the techniques involved in the invention, all of which were well known in the art even before the filing date, are highly reliable and can be readily practiced by a skilled artisan. Further, the level of skill in the art was high as of the filing date of the present application. In view of the knowledge in the art and the ample teachings of the application, one of ordinary skill in the art would readily know how to practice the claimed invention, *without undue experimentation*. Applicants respectfully request that the Examiner reconsider and withdraw the enablement rejection.

Claim Rejections under 35 U.S.C. § 102(b)

Claims 1-2, 8, 17, and 22-23 are rejected under 35 U.S.C. § 102(b) as allegedly anticipated by Heyninck et al. (Mol Cell Biol Res Commun, 2001, 4:259-65). Applicants respectfully traverse the rejections.

The standard for anticipating a claim is clearly outlined in MPEP 2131, and this standard is further supported by the courts. "A claim is anticipated only if each and every element as set forth in the claim is found, either expressly or inherently described, in a single prior art reference." *Verdegaal Bros. v. Union Oil Co. of California*, 814 F.2d 628, 631 (Fed. Cir. 1978). Applicants contend that Heyninck et al. fail to satisfy the criteria for anticipating the present invention.

Solely to expedite prosecution of the application, Applicants have amended independent claim 1 to clarify that the protein complex comprises at least TNF-α, TNFR, and TRCP1 (the elected species), or their functional variants. Dependent claim 3 has been amended to clarify that the protein complex further comprises at least one polypeptide selected from RasGAP3, NAK, and TRCP2, or their functional variants. Applicants have also amended independent claim 17 to specify that the protein complex comprises TNFR, TRCP1 (the elected species), RasGAP3, NAK, and TRCP2, or their functional variants.

Heyninck et al. (a review article) merely describe the interaction between TNFR1 and TRAF2, the interaction between TNFR and TRADD, and the interaction between TRAF2 and NAK (referred to therein as NIK) (e.g., page 259, right column, lines 14-20, 23-26; page 260, right column, lines 8-11; and Figures 1-2). However, Heyninck et al. do **not** teach an isolated, purified, or recombinant protein complex as recited in independent claim 1 or 17.

In particular, Applicants point out that Heyninck et al. do **not** teach that TRCP1, the elected species, is present in a protein complex comprising TNF-α and TNFR. The Examiner has not provided any adequate reasoning why Heyninck et al. anticipate original claims 1 and 17 with respect to TRCP1 (the elected species).

Since Heyninck et al. do not expressly or inherently teach that TRCP1 (the elected species) is present in a protein complex comprising TNFR, Heyninck et al. do not teach all the elements of independent claim 1 or 17. For the same reasons, Applicants submit that all claims depending from claim 1 or 17 are not anticipated by Heyninck et al. Applicants respectfully request that the Examiner reconsider and withdraw this rejection.

Claim Rejections under 35 U.S.C. § 103(a)

Claims 11, 12, and 25 are rejected under 35 U.S.C. § 103(a) as being unpatentable over Heyninck et al. in view of Einhauer et al. (J Biochem Biophys Methods, 2001, 49:455-65). Applicants respectfully traverse this rejection.

As described above, Heyninck et al. fail to teach all elements of independent claim 1 or 17, including at least the presence of TRCP1 in the claimed protein complex. The other cited reference (Einhauer et al.) merely teaches that the FLAG peptide may be used as a versatile fusion tag for purifying recombinant proteins (see the abstract). Einhauer et al. are totally silent on the TNF-α/TNFR signaling pathways, let alone a protein complex which comprises at least TNFR and TRCP1. Thus, Einhauer et al. fail to bridge the gap between Heyninck et al. and the claimed invention. Since the alleged combination of Heyninck et al. and Einhauer et al. fails to teach all elements of independent claim 1 or 17, all claims depending from claim 1 or 17, including claims 11, 12, and 25 which are rejected under 103(a), are not obvious over the cited references.

The Examiner has not satisfied the requirement of establishing a *prima facie* case of obviousness against independent claim 1 or 17. According to the Examination Guidelines for Determining Obviousness Under 35 U.S.C. 103 In View of the Supreme Court Decision in *KSR International Co. v. Teleflex Inc.* (Federal Register Vol. 72, No. 195 at pages 57,526-57,535) (effective October 10, 2007) ("the Guidelines"), a § 103 claim rejection based on a purported teaching, suggestion or motivation to combine prior art references to arrive at the claimed invention must support a conclusion of obviousness by including: (1) a finding that there was some teaching, suggestion or motivation to modify or combine the cited references; (2) a finding that there was a reasonable expectation of success; and (3) whatever additional findings based on the *Graham* factual inquiries may be necessary in view of the specific facts.

Applicants submit that there is no suggestion or motivation for a skilled artisan to make a protein complex comprising at least TNFR and TRCP1 as recited in claims 1 and 17. Heyninck et al. **fail** to suggest or teach the presence of any other polypeptides (besides those disclosed) which may associate with TNFR, let alone TRCP1. Even if a skilled artisan would have been motivated to identify new components in the protein complex comprising TNFR, there is no reasonable expectation of success in making the claimed protein complex. Because of the unpredictable nature of TNF-α/TNFR signaling pathways and the lack of evidence on the association between TNFR and TRCP1, a skilled artisan could not predict that a protein complex comprising at least TNFR and TRCP1 would be successfully made.

Accordingly, all claims (including claims 11, 12, and 25) are not obvious over the cited references. Applicants respectfully request reconsideration and withdrawal of the rejection under 35 U.S.C. 103(a).

CONCLUSION

In view of the above, Applicants believe that the pending application is in condition for allowance. Early and favorable reconsideration is respectfully solicited. A Petition for a one-month extension of time and appropriate fees are concurrently filed. If an additional fee is due, please charge our Deposit Account No. 18-1945, under Order No. **WYTH-P01-001** from which the undersigned is authorized to draw.

Dated: October 10, 2008

Respectfully submitted,

By _____

Z. Angela Guo
Registration No.: 54,144
ROPES & GRAY LLP
One International Place
Boston, Massachusetts 02110-2624
(617) 951-7000
(617) 951-7050 (Fax)
Attorneys/Agents For Applicant

Exhibit A

Research Articles

Combinatorial Cassette Mutagenesis as a Probe of the Informational Content of Protein Sequences

JOHN F. REIDHAAR-OLSON AND ROBERT T. SAUER

A method of combinatorial cassette mutagenesis was designed to readily determine the informational content of individual residues in protein sequences. The technique consists of simultaneously randomizing two or three positions by oligonucleotide cassette mutagenesis, selecting for functional protein, and then sequencing to determine the spectrum of allowable substitutions at each position. Repeated application of this method to the dimer interface of the DNA-binding domain of λ repressor reveals that the number and type of substitutions allowed at each position are extremely variable. At some positions only one or two residues are functionally acceptable; at other positions a wide range of residues and residue types are tolerated. The number of substitutions allowed at each position roughly correlates with the solvent accessibility of the wild-type side chain.

IT HAS BEEN MORE THAN 20 YEARS SINCE ANFINSEN AND HIS colleagues showed that the sequence of a protein contains all of the information necessary to specify the three-dimensional structure (*1*). However, the general problem of predicting protein structure from sequence remains unsolved. Part of the difficulty may stem from the complexity of protein structures. Although some 200 protein structures are known, no rules have emerged that allow structure to be related to sequence in any simple fashion (*2*). The problem is further complicated by the nonuniformity of the structural information encoded in protein sequences. Some residue positions are important, and changes at these positions can tip the balance between folding and unfolding (*3–7*). Other residues are relatively unimportant in a structural sense and a wide range of substitutions or modifications can be tolerated at these positions (*3, 7–9*).

If only a fraction of the residues in a protein sequence contribute significantly to the stability of the folded structure, then it becomes important to be able to identify these residues. We now describe the results of genetic studies that allow the importance of individual residues in protein sequences to be rapidly determined. Specifically, we determine the spectrum of functionally acceptable substitutions at residue positions near the dimer interface of the NH_2-terminal domain of phage lambda (λ) repressor (*10*). The NH_2-terminal domain binds to operator DNA as a dimer, with dimerization mediated by hydrophobic packing of α helix 5 of one monomer against α helix 5′ of the other monomer (*11*) (Fig. 1, A and B). Without helix 5 there are no contacts between the subunits (Fig. 1C). By applying combinatorial cassette mutagenesis to the helix 5 region, we find that the number and spectrum of allowable substitutions within helix 5 are extremely variable from residue to residue. In most cases, this variability can be rationalized in terms of the fractional solvent accessibility of the wild-type side chain.

General strategy. For our studies, we used a plasmid-borne gene that encodes a functional, operator-binding fragment (residues 1–102) of λ repressor (*12*). The binding of the 1–102 fragment to operator DNA depends on dimerization which, in turn, depends on the helix 5–helix 5′ packing interactions (*11, 13*). Thus, if a 1–102 protein retains normal operator-binding properties, we can infer that it is able to dimerize normally.

Mutagenesis of the helix 5 region was performed by a combinatorial cassette procedure. One example of this method, in which codons 85 and 88 are mutagenized, is illustrated in Fig. 2. On the top strand, the mutagenized codons are synthesized with equal mixtures of all four bases in the first two codon positions and an equal mixture of G and C in the third position. The resulting population of base combinations will include codons for each of the 20 naturally occurring amino acids at each of the mutagenized residue positions. On the bottom strand, inosine is inserted at each randomized position because it is able to pair with each of the four conventional bases (*14*). The two strands are then annealed and the mutagenic cassette is ligated into a purified plasmid backbone.

To identify plasmids encoding functional protein, we selected transformants for plasmid-encoded resistance to ampicillin and for resistance to killing by cI^- derivatives of phage λ. The latter selection requires that the cell express 1–102 protein that is active in operator binding (*15*). For each mutagenesis experiment, many independent transformants were chosen, single-stranded plasmid DNA was purified, and the relevant region of the 1–102 gene was sequenced. The resulting set of sequences provides a list of functionally acceptable helix 5 residues.

Substitutions in the helix 5 region. In separate experiments with different mutagenic cassettes, the codons for helix 5 residues 85 and 88; 86 and 89; 90 and 91; 84, 87, and 88; and 84, 87, and 91 were mutagenized, and genes encoding active 1–102 proteins were selected. In some cases, the survival frequency was low. For example, only 17 of 60,000 transformants passed the selection after randomization of codons 84, 87, and 88. In this case, each active candidate was sequenced. By contrast, 1,200 of 50,000 transformants passed the selection in the mutagenesis of positions 86 and 89 (*16*). In this case, we picked 50 candidates for sequence analysis. Overall, 150 active genes were sequenced (Table 1). In addition, we sequenced

The authors are in the Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139.

approximately 40 genes that had been mutagenized, but not subjected to a functional selection. These serve as controls for the efficiency of mutagenesis and also provide examples of helix 5 mutations that result in inactive 1–102 proteins (Table 1).

Many of the active sequences contain at least two residue changes compared to wild type. In principle, some of these changes could be compensatory; for example, residue X might be functionally allowed at position 85 only in combination with residue Z at position 88. This cannot be generally true, however, because most residue changes at one position were recovered in combination with several different changes at the other position or positions. It is therefore likely that most substitutions that are functionally acceptable in multiply mutant backgrounds would also be allowed as single substitutions. In Fig. 3, we show the spectrum of functionally acceptable substitutions at residue positions 84 to 91.

From the list of allowed substitutions, several conclusions may be drawn concerning the structural requirements at various positions in helix 5. We now consider these residue positions in order of decreasing "informational content," where this term is roughly defined as a value that decreases as the number of allowed substitutions increases. Thus, the informational content of a residue position is highest if only the wild-type amino acid is allowed and is lowest if each of the 20 naturally occurring amino acids is allowed.

Positions 84 and 87 in particular stand out as having a high informational content. Ile appears to be the only acceptable residue at position 84. Both Met and Leu are residues of similar size and hydrophobicity, and are the only two residues that appear to be functional at position 87. The side chains of Ile^{84} and Met^{87} form a major part of the helix-helix packing interaction at the dimer interface, where Ile^{84} of one subunit packs against $Met^{87'}$ of the other subunit, and vice versa (Fig. 4). This cluster of four residues also contacts the globular portions of the domain. Solvent accessibility calculations by the method of Lee and Richards (*17*) show that the Ile^{84} and Met^{87} side chains are almost completely buried (92 to 98 percent solvent inaccessible) in the structure of the dimer. We assume that replacement of Ile^{84} or Met^{87} with smaller side chains would diminish dimerization because hydrophobic and van der Waals interactions would be lost. In fact, mutant repressors containing Ser^{84} or Thr^{87} are defective in dimerization (*13, 18*). Replacing Ile^{84} or Met^{87} with larger residues would also be expected to be detrimental because substantial structural rearrangements would be required to accommodate larger side chains.

Seven residues (Leu, Ile, Val, Thr, Cys, Ser, and Ala) are functionally acceptable at position 91. Aromatic residues, charged residues, and strongly hydrophilic residues are not found. The wild-type Val side chain is partially buried in the dimer structure, with the Cγ2 methyl group packing against the Cδ1 methyl group of the $Ile^{84'}$ side chain. Although some of the acceptable substitutions such as Ile and Thr could make equivalent packing contacts, others such as Ala and Ser could not.

Nine residues (Trp, His, Met, Gln, Leu, Val, Ser, Gly, and Ala) are acceptable at position 90. There is a surprisingly large range in both the acceptable size and hydrophilicity of these side chains. This is especially true as the Cβ methyl group of the wild-type Ala is almost completely buried in the structure of the dimer and, at first glance, it would appear that larger side chains could not be accommodated. However, the inaccessibility of the Cβ methyl group of Ala^{90} is largely caused by the $Lys^{67'}$ side chain, which packs against it. By rotating the $Lys^{67'}$ side chain away, we were able to introduce a Trp^{90} side chain by model-building without steric clashes. Rotation of the $Lys^{67'}$ side chain away from Ala^{90} should not be energetically costly and, in fact, is observed in crystals of the NH_2-terminal domain bound to operator DNA (*19*).

Nine different residues (Trp, Tyr, Phe, Met, Ile, Val, Cys, Ser, and Ala) are functionally acceptable at position 88. There are large variations in the sizes and volumes of the acceptable side chains, although most are relatively hydrophobic. Charged residues and other strongly hydrophilic residues are not observed. In the wild-type dimer (*11*), the aromatic ring of Tyr^{88} stacks against the ring of $Tyr^{88'}$. The side chains of Trp, Phe, Met, Ile, and Val could probably form some type of packing interaction at this position, although those of Ala and Ser could not. It is known that the presence of Cys at position 88 allows a stable Cys^{88}–$Cys^{88'}$ disulfide bond, which links the monomers in a conformation that is active in operator binding (*20*).

Positions 85, 86, and 89 show considerable variability. At each of these positions, 13 different amino acids were found to function. At positions 85 and 86, aromatic, hydrophobic, polar, and charged residues are all acceptable. At position 89, aromatic residues were not represented, but each of the remaining classes was observed. In

Table 1. Sequences for the helix 5 region of active and inactive mutants obtained by combinatorial cassette mutagenesis. Active mutants are resistant to phage λKH54; these are grouped by cassette, with the wild-type sequence at the top of each group and randomized positions in boldface. Asterisks indicate sequences of mutants obtained in the absence of a functional selection. The activity of these mutants was subsequently determined by a screen. Numbers next to sequences indicate the number of times particular mutant sequences were obtained. Numbers at the tops of the columns indicate amino acid positions. The one-letter abbreviations for the amino acids are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

Active

85 90	85 90	85 90	85 90
IYEMYEAV	-L--V---	--Y--S--	--E--L--*
I--MF--- 2	-L--M---	--Y--T--	--E--E--
I--MY--- 4	-I--F---	--W--A--	--E--G--*
I-AMA---	-I--A---	--W--C--	--G--E--
I-DMY---	-I--S---	--W--Q--	
I--MA--- 3	-Q--Y---	--W--S-- 3	IYEMYEAV
I--MI---	-S--Y---	--W--K--*	------WA 4
I--LF---	-S--W---	--W--D--	------WL*
I--LW---	-T--Y---	--W--D--*	------WS
	-T--W---	--W--E--	------HV
IYEMYEAV	-T--A---	--W--G-- 3	------AV 6
I--M---V	-R--F---	--A--E-- 3	------AC 2
I--M---T	-R--W--- 4	-SM--A--	------AL 4
I--L---T	-R--A--- 3	--M--S-- 3	------AI
	-R--C---	--L--S-- 2	------AT 2
IYEMYEAV	-E--F---	--L--T--	------VA
-Y--F---	-G--Y---	--L--E--	------VC 2
-W--W--- 2		--L--E--*	------MA
-W--A---	IYEMYEAV	--I--S--	------LA*
-A--Y---	--F--M--	--Q--S--	------QV
-V--Y--- 2	--F--L--	--S--Q--	------QT
-V--A--- 3	--F--I--	--S--E--	------SV 3
-C--F--- 2	--F--Q--	--T--E--	------SC 2
-C--A---	--F--S-- 3	--D--L--	------SL 5
-L--F---	--F--T--	--D--Q--	------ST
-L--W---	--F--R--	--D--D--*	------GC
-L--A---	--F--E--	--E--L-- 2	------GI
			------GT

Inactive

85 90	85 90	85 90	85 90
A--VA---*	P--DS---*	Q--CS---*	------RN*
P--PL---*	R--TR---*	A--TP---*	------EA*
P--TN---*	T--TV---*	S--TK---*	------KV*
R--NP---*	R--VI---*	T--LN---*	------VM*
P--LL---*	L--PL---*	A--SL---*	------PA*
A--IL---*	I--LL---*	R--WS---*	------NQ*
T--KP---*	K-AIV---*	------PR*	------ME*
Q--RV---*	C--YT---*	------PP*5	------AY*
H-DVR---*			

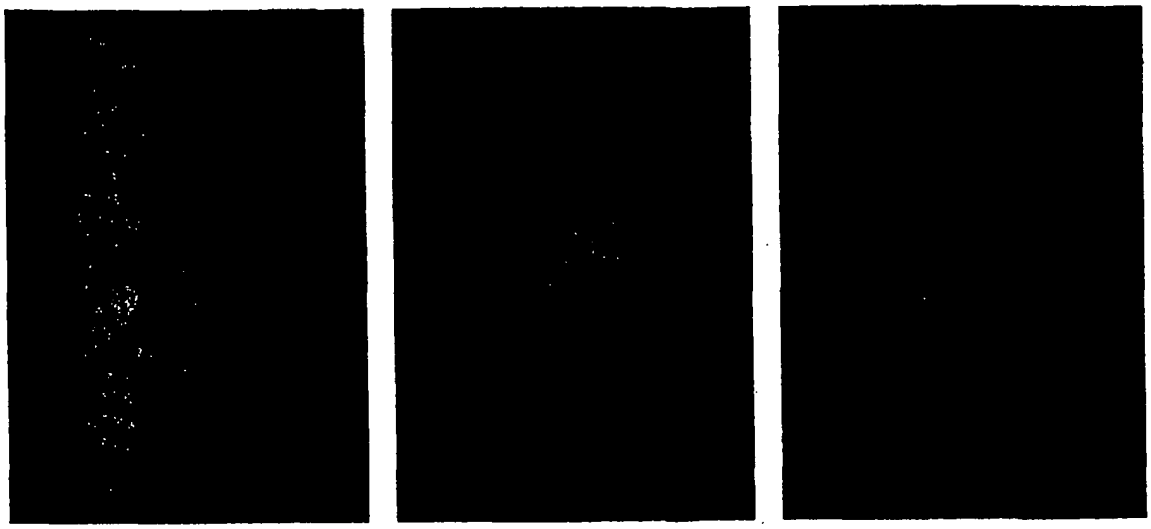

Fig. 1. Three views of the DNA-binding domain of λ repressor, showing the role of helix 5 in dimerization. (**A**) Proposed complex of repressor dimer with operator DNA (*11*). Helix 5 of each monomer is colored more lightly than the globular portion of that monomer. (**B**) Free repressor dimer, rotated 90° from the view in (A), to show the "back side" of the molecule. (**C**) Dimer with helix 5 of each monomer removed. This view illustrates the role helix 5 plays in mediating dimerization (*26*).

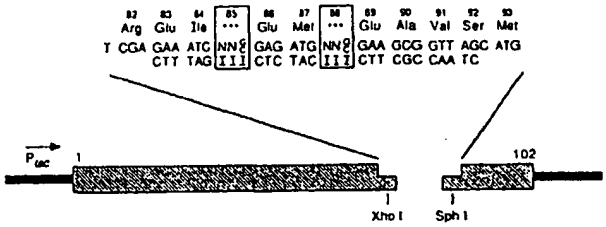

Fig. 2. Schematic diagram showing the combinatorial cassette mutagenesis procedure. At positions indicated as N, an equal mixture of A, G, C, and T was used during oligonucleotide synthesis. At positions indicated as I, inosine was used. After synthesis, the oligonucleotides were phosphorylated, annealed, and ligated into the Xho I–Sph I backbone of plasmid pJO103. Plasmid pJO103 is an M13 origin plasmid with the I–102 gene under control of a *tac* promoter; the region of the I–102 gene encoding residues 82–93 (the small Xho I–Sph I fragment) is replaced by an unrelated 1.9-kb Xho I–Sph I "stuffer" fragment. Ligated DNA was transformed into *Escherichia coli* strain X90 F′*lacI*^{Q} cells (*27*), and ampicillin-resistant colonies were selected in the presence or absence of phage λKH54. Candidates that survived the selection were cross-streaked against a series of virulent derivatives of phage λ to confirm their immunity properties [strains and methods are described in (*21*)]. Single-stranded plasmid DNA was purified from an M13RV1 transducing lysate as described (*28*), and DNA sequences were determined by the dideoxy method (*29*).

the wild-type dimer, the side chains of Tyr^{85}, Glu^{86}, and Glu^{89} are relatively solvent accessible.

Several amino acids are significantly underrepresented among the active sequences. For example, Pro is never found. This cannot be an artifact of our mutagenesis procedure because Pro is frequently observed among the unselected mutant sequences (Table 1). We conclude that Pro is not found among the functional sequences because it is selected against; its presence would presumably disrupt the α-helical structure and thereby the helix-helix packing at the dimer interface.

His, Asn, and Lys are also underrepresented among the functional helix 5 sequences. These residues are presumably not acceptable at positions 84 and 87, where the informational content is extremely high, and may not be acceptable at positions 88 and 91, where the functional substitutions are generally hydrophobic in character. The acceptability of these residues at positions such as 85 and 86 is difficult to assess from our experiments because the codons for these residues are present at reasonably low frequencies even among the unselected sequences. In these cases, we probably have not sequenced a large enough number of candidates to be confident that all acceptable substitutions have been identified. In fact, data from reversion studies (*21*) and suppressed amber studies (*22*) show that His^{85} and Lys^{86} are acceptable substitutions in the context of the intact λ repressor molecule.

Informational content and protein structure. We have com-

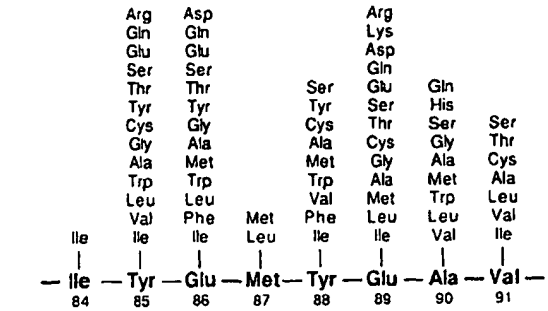

Fig. 3. Functionally acceptable residues in the helix 5 region. The amino acids are listed from top to bottom in order of increasing hydrophobicity according to the scale of Eisenberg *et al.* (*30*).

bined an efficient combinatorial mutagenesis procedure and a functional selection to probe the informational content of the eight residues that form the major part of the dimerization interface of the NH_2-terminal, operator-binding domain of λ repressor. At two of these eight residue positions, the functionally acceptable choices are highly restricted. For example, we analyzed 17 functional genes in which codon 84 had been randomized and recovered the wild-type residue, Ile, in every case. This is clearly a position of high

Fig. 4. Helix 5 residues high in informational content. The two isolated helix 5 regions of the protein are shown in green and blue. Ile^{84} and Met^{87} from the green helix are shown in yellow; Ile^{84} and Met^{87} from the blue helix are shown in red.

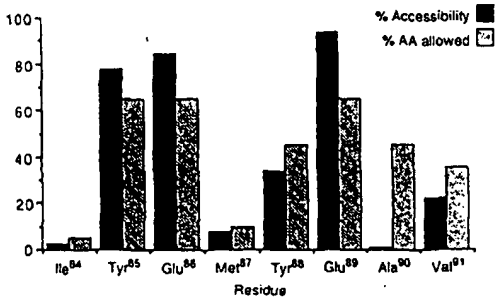

Fig. 5. Correlation between the solvent accessibility and the number of functionally acceptable substitutions. Hatched bars indicate the percentage of the 20 naturally occurring amino acids that are functionally acceptable at a residue position. Black bars indicate the fractional solvent accessibility of the wild-type side chain in the dimer. Solvent accessibilities for the NH_2-terminal domain dimer (*11*) were computed using a 1.4 Å probe by the method of Lee and Richards (*17*). Fractional accessibilities were obtained by dividing by the appropriate side chain accessibilities calculated for the monomer. The fractional accessibilities change only slightly if the side chain accessibilities in the reference tripeptide Ala-X-Ala (*17*) are used instead as the reference state.

informational content. The informational content is also high at position 87, where Met and Leu are the only acceptable residues. By contrast, the remaining positions have moderate to low informational contents. For example, among 38 functional genes in which codon 85 had been randomized, the wild-type residue was recovered only once, and 12 other residues, differing in size and chemical properties, were recovered in the remaining cases. This is clearly a position of low informational content. It is striking that most of the structural determinants of dimerization in this eight-residue segment reside in two residues only. The remaining positions are surprisingly tolerant of a wide range of substitutions. If this high level of tolerance is generally true of protein sequences, then the problem of understanding and predicting structure may rest largely on the ability to identify those few residues that are crucial.

The positional variability of the informational content in helix 5 can, in general, be rationlized in terms of the solvent accessibility of the wild-type residues in the crystal structure (*11*). There is a rough correlation between the number of acceptable substitutions and the fractional extent to which the wild-type side chain is solvent accessible (Fig. 5). At exposed surface positions such as 85, 86, and 89, we find that many different residues and residue types can be functionally accommodated. By contrast, at positions such as 84 and 87, where the wild-type side chain is almost completely buried, we find that the functionally acceptable residue choices are extremely restricted. There is one apparent exception to the simple rule that buried residues are high in informational content. Ala^{90} is inaccessible to solvent in the crystal structure, and yet we find that many substitutions are allowed at this position. However, the inacessibility of the Ala^{90} side chain to solvent is not due to close packing at the dimer interface, but rather to an interaction with a nearby surface side chain. This side chain can presumably move to allow larger side chains to be accommodated at position 90. Examples of this type demonstrate the need to distinguish between two types of buried side chains: those that can become exposed by relatively minor rearrangement of other side chains, and those that are tightly packed in the hydrophobic core.

There is no reason to assume that there should always be a strict correlation between the solvent accessibility of a residue and the structural informational content of that position. For one thing, the chemical properties of the 20 amino acids are not related in any simple linear fashion. Moreover, the structural importance of some residues in proteins almost certainly stems from interactions other than simple hydrophobic packing. Nevertheless, the closely packed nature of protein interiors (*23*) provides a simple molecular explanation for the structural importance of buried residues, and destabilizing mutations are commonly found to affect hydrophobic core residues (*3*–*7*). By contrast, missense mutations or chemical modifications that affect surface residues are often found to have little or no influence on protein stability (*3*, *7*, *8*). Thus, it is reasonable that solvent accessibility should be an extremely important determinant of the informational content of a residue position.

Our overall strategy for rapidly probing informational content should be broadly applicable to a wide range of protein structure-function problems in systems where genetic selections or screens can be devised. The method consists of three basic elements: (i) the use of cassette mutagenesis to introduce extremely high levels of targeted random mutagenesis; (ii) the use of a functional selection to identify genes encoding active proteins; and (iii) the use of rapid DNA sequencing methods to determine the spectrum of functionally acceptable residues in a relatively large number of candidates. Our method of combinatorial cassette mutagenesis (Fig. 2) allows several residue positions to be mutagenized at the same time and, in principle, generates a mutant population in which each of the 20 amino acids is represented at each mutagenized position (*24*). When two or three codons are mutagenized at the same time, the entire analysis is able to proceed more rapidly. Moreover, at this level of mutagenesis most two-residue and three-residue combinations should be present in the mutagenized population and should be recovered if they result in a functional protein. In our study of the packing of the 84 and 87 side chains, we recovered only two (Ile^{84} with Met^{87} and Ile^{84} with Leu^{87}) of the 400 possible residue combinations. Thus, because both positions were mutagenized in the same experiment, we are able to conclude that there are not significantly different ways of packing the dimer interface.

In principle, data like that shown in Fig. 3 could be generated for an entire protein sequence, and additional experiments could be devised to determine whether the positions of high informational content were important for structure or function. For proteins of unknown structure, such data might be quite useful for structural predictions. First, current predictive algorithms could be applied to the family of related sequences generated by our method, as each of these sequences is able to form the same basic structure. Second, because of their fundamental repeats, α-helical and β-strand regions might be recognized by characteristic patterns of high and low informational content. Third, the positions of highest structural informational content should include the residues involved in

formation of the hydrophobic core of the protein. This information might prove useful in combination with the tertiary template ideas recently proposed (*25*).

REFERENCES AND NOTES

1. C. J. Epstein, R. F. Goldberger, C. B. Anfinsen, *Cold Spring Harbor Symp. Quant. Biol.* **28**, 439 (1963); C. B. Anfinsen, *Science* **181**, 223 (1973).
2. T. E. Creighton, *Proteins: Structures and Molecular Properties* (Freeman, New York, 1983), chap. 6.
3. M. H. Hecht, H. C. M. Nelson, R. T. Sauer, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 2676 (1983); M. H. Hecht, J. M. Sturtevant, R. T. Sauer, *ibid.* **81**, 5685 (1984); M. H. Hecht, K. M. Hehir, H. C. M. Nelson, J. M. Sturtevant, R. T. Sauer, *J. Cell. Biochem.* **29**, 217 (1985).
4. D. Shortle and B. Lin, *Genetics* **110**, 539 (1985); D. Shortle and A. K. Meeker, *Proteins* **1**, 81 (1986).
5. A. Pakula, V. Young, R. T. Sauer, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 8829 (1986).
6. T. Alber, S. Dao-pin, J. A. Nye, D. C. Muchmore, B. W. Matthews, *Biochemistry* **26**, 3754 (1987).
7. G. Fermi and M. F. Perutz, *Haemoglobin and Myoglobin* (Clarendon, Oxford, 1981).
8. M. Hollacker and T. E. Creighton, *Biochem. Biophys. Acta* **701**, 395 (1982).
9. J. H. Miller, in *The Operon*, J. H. Miller and W. S. Reznikoff, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1978), pp. 31–88.
10. λ repressor consists of an NH_2-terminal and COOH-terminal domain; each domain contains an independent set of dimerization contacts [C. O. Pabo, R. T. Sauer, J. M. Sturtevant, M. Ptashne, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 1608 (1979); C. O. Pabo, thesis, Harvard University (1980)].
11. C. O. Pabo and M. Lewis, *Nature* **298**, 443 (1982).
12. The NH_2-terminal domain of λ repressor is minimally comprised of residues 1–92. This is the fragment for which the crystal structure is known. In our studies we use a slightly larger NH_2-terminal fragment consisting of repressor residues 1–102. Nuclear magnetic resonance studies show that the 1–92 and 1–102 fragments have the same basic structure and similar dimerization properties (*13*). The operator binding properties of the 1–92 and 1–102 fragments are also extremely similar (*20*).
13. M. A. Weiss, C. O. Pabo, M. Karplus, R. T. Sauer, *Biochemistry* **26**, 897 (1987); M. A. Weiss, M. Karplus, R. T. Sauer, *ibid.*, p. 890.
14. F. H. Martin, M. M. Castro, F. Aboul-ela, I. Tinoco, *Nucleic Acids Res.* **13**, 8927 (1985).
15. In control experiments, we found that cells containing at least 5 to 10 percent of the wild-type activity could survive the phage selection. This figure was determined by Western analysis of lysates of cells containing the wild-type 1–102 gene under control of the inducible *lac* promoter. When expression of 1–102 was induced to a level sufficient to make cells resistant to phage λKH54, the intracellular level of 1–102 was 5 to 10 percent of that produced from the *tac* promoter.
16. The survival frequency can be somewhat misleading, as some cells containing functional genes do not survive the selection. For example, by screening the unselected candidates in the mutagenesis of positions 86 and 89, we found that approximately 20 percent of the cells contained active protein. By contrast, only 2.4 percent of the cells survived the selection in this experiment.
17. B. Lee and F. M. Richards, *J. Mol. Biol.* **55**, 379 (1971); in our studies, accessible surface areas were calculated with the use of the ACCESS program written by T. Richmond.
18. F. S. Gimble and R. T. Sauer, *J. Mol. Biol.*, in press.
19. S. Jordan and C. Pabo, personal communication.
20. R. T. Sauer *et al.*, *Biochemistry* **25**, 5992 (1986); C. O. Pabo and E. G. Suchanek, *ibid.* **25**, 5987 (1986).
21. M. H. Hecht and R. T. Sauer, *J. Mol. Biol.* **186**, 53 (1985).
22. H. C. M. Nelson, M. H. Hecht, R. T. Sauer, *Cold Spring Harbor Symp. Quant. Biol.* **47**, 441 (1983).
23. F. M. Richards, *J. Mol. Biol.* **82**, 1 (1974); C. Chothia, *Nature* **254**, 304 (1975).
24. There are several methods, in addition to our own, that could be used for combinatorial cassette mutagenesis: A. R. Oliphant, A. L. Nussbaum, K. Struhl, *Gene* **44**, 177 (1986); K. M. Derbyshire, J. J. Salvo, N. D. F. Grindley, *ibid.* **46**, 145 (1986). The use of inosines, in our method, is convenient and efficient, but does lead to some bias in the frequency at which each base is recovered. For example, among the unselected candidates in the mutagenesis of positions 84, 87, and 88, the frequencies at which the four bases were recovered at the mutagenized positions were: 23 percent A, 35 percent C, 20 percent G, and 23 percent T. The methods cited above avoid a pairing bias by performing enzymatic second-strand synthesis.
25. J. W. Ponder and F. M. Richards, *J. Mol. Biol.* **193**, 775 (1987).
26. These graphics were produced with the Promodeber molecular graphics program (New England BioGraphics). Coordinates were provided by C. O. Pabo.
27. D. Hanahan, *J. Mol. Biol.* **166**, 557 (1983).
28. A. K. Vershon, K. Blackmer, R. T. Sauer, in *Protein Engineering: Applications in Science, Medicine, and Industry*, M. Inouye and R. Sarma, Eds. (Academic Press, Orlando, FL, 1986), pp. 243–256.
29. F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977).
30. D. Eisenberg, R. M. Weiss, T. C. Terwilliger, W. Wilcox, *Faraday Symp. Chem. Soc.* **17**, 109 (1982).

31. We thank R. Breyer for providing plasmids and antibodies required for the work, D. Vershon for suggesting the use of inosines for the combinatorial mutagenesis method, and J. Bowie for pointing out the potential uses of the method in structural prediction. We also thank C. Pabo and S. Jordan for providing the coordinates of the NH_2-terminal domain and its complex with operator DNA, and thank G. Quigley for help with the accessibility calculations. This work was supported by NIH grant AI-15706 and by a predoctoral grant (to J.R.-O.) from the National Science Foundation.

3 March 1988; accepted 18 May 1988